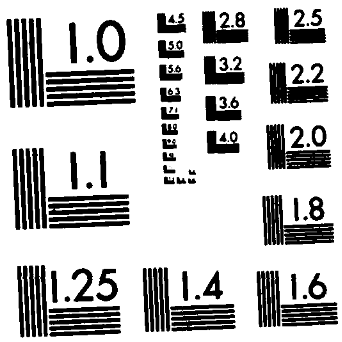
AD-A134 533 A PARAMETRIC BOOTSTRAP PROCEDURE FOR TESTING SEPARATE FAMILIES OF HYPOTHESES(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER W LOH SEP 83 MRC-TSR-2574 1/1

UNCLASSIFIED DAAG29-80-C-0041 F/G 12/1 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

AD-A134533

MRC Technical Summary Report #2574

A PARAMETRIC BOOTSTRAP PROCEDURE FOR TESTING SEPARATE FAMILIES OF HYPOTHESES

Wei-Yin Loh

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

September 1983

(Received August 10, 1983)

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

A PARAMETRIC BOOTSTRAP PROCEDURE FOR TESTING SEPARATE FAMILIES OF HYPOTHESES

Wei-Yin Loh

Technical Summary Report #2574

September 1983

ABSTRACT

It is shown that the size of the parametric bootstrap procedure of Williams (1970) is biased upward. A bias-corrected version is shown to be better. The finite-sample performance of this procedure is examined and compared with that of Cox's (1961, 1962) test in a number of examples. In some of these, the proposed test reduces to the traditional optimal test.

AMS (MOS) Subject Classification: 62F03

Key Words: Parametric bootstrap; Separate families of hypothesis; Spline; Golden-section search

Work Unit Number 4 - Statistics and Probability

Department of Statistics and Mathematics Research Center, University of Wisconsin-Madison, Madison, WI 53705.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-7927062, Mod 2 and by funds from the University of Wisconsin Graduate School Research Committee.

SIGNIFICANCE AND EXPLANATION

This paper proposes a procedure for testing non-nested families of hypotheses which substitutes raw computing power for asymptotic approximations. Given access to a modern computer, the procedure is practically universally applicable. Examples illustrating its small-sample properties are provided, as are theorems on its asymptotic behavior.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

A PARAMETRIC BOOTSTRAP PROCEDURE FOR TESTING SEPARATE FAMILIES OF HYPOTHESES

Wei-Yin Loh

1. Introduction.

This paper considers the problem of finding general procedures for testing separate families of hypotheses, separate in the sense that an arbitrary member in the null cannot be obtained as a limit of members in the alternative hypothesis. A fairly general test has been proposed in two well-known papers by Cox (1961, 1962). It is based on the property that, subject to regularity conditions, the normalized logarithm of the ratio of maximized likelihoods behaves asymptotically like a standard normal random variable.

Specifically, let $(X_1,\ldots,X_n)$ be a random sample from a distribution with density $h(x)$ and consider testing the separate families $H_0 : h(x) = f(x,\theta)$ vs. $H_1 : h(x) = g(x,\omega)$ where (θ,ω) are unknown, possibly vector-valued, parameters. Let $\hat{\theta}$, $\hat{\omega}$ be the maximum likelihood estimates of θ, ω respectively and

$$T_n = n^{-1} \Sigma \log\{g(x_i,\hat{\omega})/f(x_i,\hat{\theta})\} \quad . \tag{1.1}$$

Then, under H_0 and subject to appropriate regularity conditions (e.g. White, 1982),

$$Z \equiv n^{1/2}(T_n - E_{\hat{\theta}}T_n) \to N(0, \sigma^2(\theta)) \tag{1.2}$$

Department of Statistics and Mathematics Research Center, University of Wisconsin-Madison, Madison, WI 53705.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-7927062, Mod. 2 and by funds from the University of Wisconsin Graduate School Research Committee.

for some $\sigma^2(\theta) > 0$. In this paper $N(\mu,\sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. Under H_1, Z tends to $+\infty$ almost surely. Hence large values of Z are evidence against H_0. At nominal level α, Cox's (1961, 1962) test, denoted by ϕ_{COX}, rejects H_0 if $Z > z_{1-\alpha}\sigma(\hat{\theta})$, where z_α is the $N(0,1)$ α-quantile.

An obviously desirable property of ϕ_{COX} is that the probability of a type I error, α_I, converges to the nominal level as n increases. However, no study seems to have been done on the asymptotic behavior of the size of ϕ_{COX}.

A variant of ϕ_{COX}, to which it is asymptotically equivalent under the null hypothesis, has been suggested by Atkinson (1970). However this was shown to be not always consistent by Pereira (1977). Some authors have proposed tests based on statistics other than the likelihood ratio. Epps, Singleton and Pulley (1982) use the empirical moment generating function, and Shen (1982) and Sawyer (1983) derive their tests from information theoretic considerations. None of these tests appear to be superior to the others.

A common feature of the tests is that they all depend on some form of asymptotic approximation, and so require various regularity conditions for their validity. In an attempt to obtain a solution to a problem where such conditions are absent, Williams (1970a,b) considers a different approach which substitutes raw computing power for asymptotics. Given the data, Williams (1970) proposes simulating the distribution of T_n in (1.1) on a computer assuming that $\theta = \hat{\theta}$. The null hypothesis is rejected at the nominal α level if the observed

value of T_n is greater than the $(1-\alpha)$-quantile of the simulated distribution. We will call this procedure ϕ_{PAR} since it has been named "parametric bootstrap" in other contexts by some authors (e.g. Efron, 1982).

It seems pertinent to remark here that although the idea of using computer simulation to obtain the critical values of a test statistic is not widely practised at the present time, there are problems where no better alternative exists. For example, when the hypotheses represent location-scale families of distributions, the well-known uniformly most powerful invariant test statistic is a ratio of multiple integrals whose null distribution, exact or approximate, is unknown. However it can be approximated to any desired degree of accuracy by Monte Carlo simulation quite easily, especially as the null distribution is independent of the unknown parameters.

One aim in this paper is to compare the finite-sample performance of ϕ_{COX} and ϕ_{PAR}, with emphasis on (i) the extent to which the size of the test, α_I^s, exceeds the nominal α, and (ii) the power of the tests. It is shown in section 2 that $\alpha_I^s(\phi_{PAR})$ is never less than α. In fact, an example is given in section 3 where $\alpha_I^s(\phi_{PAR}) = 1$ for all α and n. We also show that $\alpha_I^s(\phi_{COX})$ can be either biased upward or downward.

To reduce the bias in $\alpha_I^s(\phi_{PAR})$, a simple modification, ϕ_*, is proposed in section 2. It is shown that if under H_0, $\hat{\theta}$ is a consistent estimator and the $1 - \alpha$ quantile of T_n is a sufficiently smooth function of θ for each n, then $\alpha_I^s(\phi_*)$ is bounded above by

$\alpha + \epsilon_n$ for some $\epsilon_n > 0$ which tends to zero as $n \to \infty$. No explicit conditions on the limiting distributional behavior of T_n are assumed. Sections 3 and 4 contain examples comparing the different tests.

2. Improving on ϕ_{PAR}.

Let T_n be given in (1.1). For each θ, α and n, define the critical value $c(\theta,\alpha,n)$ such that

$$P_\theta(T_n > c(\theta,\alpha,n)) = \alpha \quad . \tag{2.1}$$

We assume for simplicity here that T_n is a continuous random variable. Let $c_m(\alpha,n) = \sup_\theta c(\theta,\alpha,n)$. If $c_m(\alpha,n)$ is known, the test ϕ_{LRT} which rejects H_0 if $T_n > c_m(\alpha,n)$ is clearly level α. In fact, under mild regularity assumptions, ϕ_{LRT} is asymptotically optimal in terms of Bahadur efficiency (see e.g. Bahadur, 1971 or Brown, 1971). This does not, of course, imply that ϕ_{LRT} is necessarily most powerful level α for finite n.

While $c(\theta,\alpha,n)$ can be approximated given θ, α and n, by brute-force computer simulation if necessary, the computation of $c_m(\alpha,n)$ presents a much harder problem. The parametric bootstrap ϕ_{PAR} avoids this by having the rejection region: $T_n > c(\hat{\theta},\alpha,n)$. Notice that since $c(\hat{\theta},\alpha,n) < c_m(\alpha,n)$, ϕ_{PAR} is at least as powerful as ϕ_{LRT}. Our first theorem shows that this is obtained at a cost.

__Theorem 2.1.__ For all α and n,

$$\alpha_I^s(\phi_{PAR}) = \sup_\theta P_\theta(T_n > c(\hat{\theta},\alpha,n)) > \alpha \quad .$$

__Proof.__ $\sup_\theta P_\theta(T_n > c(\hat{\theta},\alpha,n)) > \sup_\theta P_\theta(T_n > c_m(\alpha,n)) = \alpha$.

The bias in the size of ϕ_{PAR} will be reduced if we use a critical value for T_n that is greater than $c(\hat{\theta},\alpha,n)$. Assuming that $\hat{\theta}$ is a consistent estimator of θ, this can be done in the following way. Let $I_n(\hat{\theta})$ be a $100(1-\alpha_n)\%$ confidence interval for θ such that both its

length and α_n tend to zero as n tends to infinity. The idea behind the test we will propose is to use $c^* = c(\theta^*,\alpha,n)$ as the critical value of T_n, where θ^* maximizes $c(\theta,\alpha,n)$ over $I_n(\hat{\theta})$. Of course, this method is practicable only if θ^* is known a priori for all $I_n(\hat{\theta})$. Section 3 contains examples where this is the case. For other cases, θ^* can at best be approximated with an "interval of uncertainty" (θ_1^*,θ_2^*), which will contain θ^* if it is assumed that $c(\theta,\alpha,n)$ is unimodal in $I_n(\hat{\theta})$. Making this assumption, the standard optimum-seeking methods like the Fibonacci and golden-section search can be used. Given $\varepsilon' > 0$, each of these two methods is known to produce an interval of uncertainty, of length less than ε', with a minimum number of function evaluations; see e.g. Wilde (1964).

To approximate $c(\theta^*,\alpha,n)$, let $I_n(\hat{\theta}) = (\hat{\theta}_1,\hat{\theta}_2)$ and $\varepsilon = \theta_2^* - \theta_1^*$. Suppose first that $\hat{\theta}_1 < \theta_1^* < \theta_2^* < \hat{\theta}_2$. Then we may assume without loss of generality that $\hat{\theta}_1 < \theta_1^* - \varepsilon < \theta_2^* + \varepsilon < \hat{\theta}_2$, because if this were not true, we could reduce the interval of uncertainty, and hence ε, by continuing the golden-section search. Let $m_1 = c(\theta_1^*,\alpha,n) - c(\theta_1^* - \varepsilon, \alpha,n)$ and $m_2 = c(\theta_2^* + \varepsilon, \alpha,n) - c(\theta_2^*,\alpha,n)$, and define

$$c^*(\hat{\theta},\alpha,n) = \max_{i=1,2} \{c(\theta_i^*,\alpha,n) + |m_i|\} \quad . \tag{2.2}$$

If $\hat{\theta}_i = \theta_i^*$ for some $i = 1,2$, we set $m_i = 0$. Denote by ϕ_* the test which rejects the null hypothesis if $T_n > c^*(\hat{\theta},\alpha,n)$. If $c(\theta,\alpha,n)$ is sufficiently smooth, an upper bound on the size of ϕ_* can be obtained.

Theorem 2.2.

(i). $\alpha_I^s(\phi_*) < \alpha_I^s(\phi_{PAR})$.

(ii). Suppose that (a) for each local maximum $\tilde{\theta}$ of $c(\theta,\alpha,n)$ there is $\delta = \delta(\tilde{\theta}) > 0$ such that $c(\theta,\alpha,n)$ is concave for $\theta \in (\tilde{\theta}-2\delta,\tilde{\theta}+2\delta)$ and, (b) with H_0-probability one, $I_n(\hat{\theta})$ contains at most one local maximum of $c(\theta,\alpha,n)$. Then $\alpha_I^s(\phi_*) < \alpha + \alpha_n$ if whenever $\tilde{\theta}$ obtains and $\tilde{\theta} \in I_n(\hat{\theta})$, we choose ε in (2.2) so small that $\varepsilon < \delta(\tilde{\theta})$.

Proof. (i). Obvious.

(ii). The assumptions imply that for every θ,

$$\begin{aligned} P_\theta(\phi_* \text{ rejects } H_0) &< P_\theta(T_n > c(\theta,\alpha,n),\ \theta \in I_n(\hat{\theta})) \\ &\quad + P_\theta(\theta \notin I_n(\hat{\theta})) \\ &< \alpha + \alpha_n . \end{aligned}$$

Both assumptions in part (ii) of the theorem set conditions on the smoothness of $c(\theta,\alpha,n)$. Condition (b) is necessary to ensure that the search does not yield the 'wrong' local maximum. Although the local maxima $\{\tilde{\theta}\}$ are seldom known in advance, the sequential nature of the search allows the experimenter to plot the points $\{(\theta^{(j)},c(\theta^{(j)},\alpha,n)\}$ at each stage and decide for himself whether the length ε of the current interval of uncertainty is small enough for stopping. Stopping at any stage amounts to making assumption (a) of the theorem with $\delta(\tilde{\theta}) > \varepsilon$, if there exists $\tilde{\theta}$ in the observed $I_n(\hat{\theta})$.

In the above discussion it is assume that $c(\theta,\alpha,n)$ can be obtained exactly for any θ selected by the search procedure. When $c(\theta,\alpha,n)$ has to be estimated by computer simulation, its value will be subject to Monte Carlo error. However, since this error can be made arbitraily small by increasing the number of Monte Carlo replicates, we assume it to be negligible here.

3. Some examples permitting analytic solution.

In this section the superiority of ϕ_* is demonstrated in some classical testing problems where analytic solutions are possible.

Example 3.1. Testing a normal mean.

Let $(X_1,\ldots,X_n)$ be a random sample from $N(\mu,\sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$, and consider testing $H_0 : N(0,\sigma^2)$ vs. $H_1 : N(\mu,1)$. It is easily seen from (1.1) that

$$T_n = .5(1 + \log s^2 - s^2 + \bar{x}^2)$$

where $\bar{x}$ is the sample mean and s^2 the maximum likelihood estimate of σ^2. Thus ϕ_{PAR} has the rejection region

$$.5(1 + \log s^2 - s^2 + \bar{x}^2) > k_{1-\alpha}(s)$$

where $k_\alpha(\sigma)$ is the α quantile of T_n when σ obtains. We can avoid the evaluation of $k_{1-\alpha}(s)$ by rewriting the rejection region as $\bar{x}^2/s^2 > k'$. Now k', being the $1 - \alpha$ quantile of $\bar{x}^2/s^2$, is independent of s. So both ϕ_{PAR} and ϕ_* are equivalent to the t-test, which is uniformly most powerful unbiased. It turns out that ϕ_{COX} does not exist for this problem because $T_n - E_{\hat{\sigma}}T_n$ is of order n^{-1} and hence the LHS of (1.2) converges to 0 in probability under H_0.

Example 3.2. Testing a normal variance.

Let $(X_1,\ldots,X_n)$ be a random sample from $N(\mu,\sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$. We consider testing the following hypotheses.

(a) $H_0 : \sigma^2 = 1$ vs. $H_1 : \sigma^2 = 2$.

It can be verified that (1.1) gives $T_n = s^2/4 + \text{constant}$, where s^2 is the maximum likelihood estimate of σ^2. Since the distribution of T_n

is independent of μ, ϕ_{PAR} and ϕ_* are identical, and yield the uniformly most powerful test.

An easy calculation shows that at the nominal α level, ϕ_{COX} rejects H_0 if

$$ns^2 > n-1 + z_{1-\alpha}\{2(n-1)\}^{1/2} . \tag{3.1}$$

The RHS is precisely the two-term Cornish-Fisher expansion for the $(1-\alpha)$-quantile $\chi^2_{n-1;1-\alpha}$ of the χ^2_{n-1} distribution. Table 3.1 gives some numerical values of the size of the test, $\alpha^s_I(\phi_{COX})$, for different values of α and n. Numbers in parentheses give the ratio α^s_I/α. The entries indicate that $\alpha^s_I(\phi_{COX})$ is biased upward.

Table 3.1. Values of α^s_I and α^s_I/α for ϕ_{COX}

$(H_0 : \sigma^2 = 1$ vs. $H_1 : \sigma^2 = 2)$

α \ n	5	10	20	100
.05	.070(1.4)	.067(1.34)	.064(1.28)	.057(1.14)
.01	.032(3.2)	.026(2.6)	.022(2.2)	.016(1.6)
.005	.024(4.8)	.018(3.6)	.014(2.8)	.009(1.8)

(b) $H_0 : \sigma^2 = 1$ vs. $H_1 : \sigma^2 = 1/2$.

As in (a), ϕ_{PAR} and ϕ_* are both equivalent to the uniformly most powerful invariant test which rejects for small s^2. The rejection region for ϕ_{COX} is

$$ns^2 < n-1 - z_{1-\alpha}\{2(n-1)\}^{1/2} . \tag{3.2}$$

Table 3.2 shows that $\alpha_I^s(\phi_{COX})$ is now biased downward; numbers in parentheses again give α_I^s/α. The reason for the many zeros in the table is because the RHS of (3.2) is negative or close to 0 for those values of α and n.

Table 3.2. Values of α_I^s and α_I^s/α for ϕ_{COX}

$(H_0 : \sigma^2 = 1$ vs. $H_1 : \sigma^2 = .5)$

α \ n	5	10	20	100
.05	0(0)	.009(.18)	.024(.48)	.040(.8)
.01	0(0)	0(0)	0(0)	.005(.5)
.005	0(0)	0(0)	0(0)	.002(.4)

(c) $H_0 : \sigma^2 \le 1$ vs. $H_1 : \sigma^2 > 1$.

Let $\hat{\sigma}_i^2$ be the maximum likelihood estimate of σ^2 under H_i, i = 0,1. Then $\hat{\sigma}_0^2 = \min(s^2,1)$, $\hat{\sigma}_1^2 = \max(s^2,1)$ and

$$T_n = \{.5(s^2-1) - \log s\} \operatorname{sgn}(s^2-1) \quad ,$$

the latter being an increasing function of s^2. Hence ϕ_{PAR} rejects H_0 at nominal level α if $ns^2 > \hat{\sigma}_0^2\chi^2_{n-1,1-\alpha}$, which reduces to

$$ns^2 > \chi^2_{n-1,1-\alpha} \quad \text{if} \quad s^2 > 1$$

$$n > \chi^2_{n-1,1-\alpha} \quad \text{if} \quad s^2 < 1 \quad .$$

Clearly for α-values satisfying $\chi^2_{n-1,1-\alpha} > n$, ϕ_{PAR} yields the uniformly most powerful level α test. This will be the case for the levels used in practice. Otherwise, ϕ_{PAR} rejects H_0 regardless of the data. Therefore

$$\alpha_I^s(\phi_{PAR}) = \begin{cases} \alpha & \text{if } \chi^2_{n-1,1-\alpha} > n \\ 1 & \text{otherwise} \end{cases} .$$

Since for fixed $1/2 < \alpha < 1$,

$$\nu - \chi^2_{\nu,1-\alpha} = O(\nu^{-1/2}) \quad \text{as} \quad \nu \to \infty \tag{3.3}$$

we conclude that ϕ_{PAR} is not even asymptotically level α for $\alpha \in (.5,1)$.

To derive ϕ_*, let the interval with endpoints $\hat{\sigma}_0^2 \exp(\pm n^{-1/2} k_n)$ be a confidence interval for σ^2 under H_0 such that $k_n \to \infty$, and $n^{-1/2} k_n \to 0$ as $n \to \infty$. The rejection region for ϕ_* then has the form $ns^2 > \hat{\hat{\sigma}}_0^2 \chi^2_{n-1,1-\alpha}$ where $\hat{\hat{\sigma}}_0^2 = \min\{s^2 \exp(n^{-1/2} k_n), 1\}$. Hence ϕ_* rejects H_0 if

$$\{ns^2 > \chi^2_{n-1,1-\alpha} \text{ and } s^2 \exp(n^{-1/2} k_n) > 1\}$$

or

$$\{n \exp(-n^{-1/2} k_n) > \chi^2_{n-1,1-\alpha} \text{ and } s^2 \exp(n^{-1/2} k_n) < 1\} .$$

Clearly, at the usual levels of α ($< 1/2$), ϕ_* is also uniformly most powerful level α for sufficiently large n. Unlike ϕ_{PAR}, however, (3.3) and the conditions on k_n imply that $\alpha_I^s(\phi_*) \to \alpha$ as $n \to \infty$ for all $0 < \alpha < 1$.

It is noted that ϕ_{COX} is not valid in the present situation because the LHS of (1.2) does not converge to a normal distribution. This is partly due to the fact that $n^{1/2}(\hat{\sigma}_1^2 - 1)$ is not asymptotically normal (cf. White, 1982).

Similar results to (c) carry over to the problem of testing $H_0 : N(\theta,1),\ \theta \leq \theta_0$ vs. $H_1 : N(\theta,1),\ \theta > \theta_0$. The next example shows ϕ_{PAR} at its worst.

Example 3.3. Testing the location of an exponential distribution.

Let $(X_1,\ldots,X_n)$ be a random sample from the exponential distribution with density $\exp(\theta-x)$, $x > \theta$, and consider testing $H_0 : \theta > 0$ vs. $H_1 : \theta < 0$. The maximum likelihood estimators of θ under H_0 and H_1 are $\hat{\theta}_0 = X^+_{(1)}$, $\hat{\theta}_1 = X^-_{(1)}$ respectively, where $X_{(1)}$ is the smallest order statistic and $x^- = \min(x,0)$, $x^+ = \max(x,0)$. An easy calculation yields

$$T_n = \begin{cases} -X_{(1)} & \text{if } X_{(1)} > 0 \\ \infty & \text{if } X_{(1)} < 0 \end{cases} .$$

Since $P_\theta(X_{(1)} < \theta - n^{-1}\log(1-\alpha)) = \alpha$, ϕ_{PAR} rejects H_0 if $X_{(1)} < 0$ or $X_{(1)} < \hat{\theta}_0 - n^{-1}\log(1-\alpha)$. Substitution for $\hat{\theta}_0$ shows that for all values of $X_{(1)}$, at least one of these two inequalities is satisfied. Hence ϕ_{PAR} rejects H_0 with probability one for all α and n.

To derive ϕ_*, we use the fact that under H_0, $(X_{(1)} + n^{-1}\log\alpha_n, X_{(1)})$ is a $100(1-\alpha_n)\%$ confidence interval for θ, where $\alpha_n \to 0$. Using this ϕ_* has the rejection region

$$X_{(1)} < 0 \quad\text{or}\quad X_{(1)} < \{X_{(1)} + n^{-1}\log\alpha_n\}^+ - n^{-1}\log(1-\alpha) .$$

This reduces to

$$X_{(1)} < \min\{-n^{-1}\log(1-\alpha), -n^{-1}\log\alpha_n\}$$

or

$$\{X_{(1)} > -n^{-1}\log\alpha_n \text{ and } -n^{-1}\log\alpha_n < -n^{-1}\log(1-\alpha)\} .$$

Let n_α be the greatest integer n satisfying $-n^{-1}\log\alpha_n < -n^{-1}\log(1-\alpha)$. If $n < n_\alpha$, ϕ_* rejects H_0 with probability one regardless of the data. On the other hand, if $n > n_\alpha$, ϕ_* rejects H_0 whenever $X_{(1)} < -n^{-1}\log(1-\alpha)$. This coincides with the rejection

region of the uniformly most powerful level α test. Therefore we have $\alpha_I^s(\phi_*) = \alpha$ for all $n > n_\alpha$ and all α. As in part (c) of the earlier example, ϕ_{COX} is inapplicable here.

4. Three examples considered in Cox (1962).

We consider in this section three examples in Cox (1962) where analytic solutions for the level and power of the tests are not available. The comparisons are therefore based on Monte Carlo simulation. Only the nominal level of $\alpha = .05$ is investigated.

Although the procedure ϕ_* as described in section 2 is computationally quite easy to apply on any one data set, a Monte Carlo evaluation of its performance using, say, 10^4 simulated data sets can be a time consuming task. To reduce this effort, the following modification of ϕ_* is adopted here. In the first stage, after deciding on the parameter-ranges of interest, a grid of between 20 to 30 θ-values is selected. For each selected θ, the critical value $c'(\theta,\alpha,n)$ of $n^{1/2}T_n$ is approximated on a computer by simulating 10,001 values of $n^{1/2}T_n$ and setting $c'(\theta,\alpha,n)$ to be the 9,501st ordered value. The points $\{(\theta,c'(\theta,\alpha,n))\}$ are then smoothly interpolated with a cubic spline to yield an approximation $c_s(\theta,\alpha,n)$ to $c'(\theta,\alpha,n)$. This curve is stored for use in the second stage, where for each desired member of the null or alternative hypotheses, 10^4 sets of pseudo-random samples of size n are simulated. For each set, the values of $n^{1/2}T_n$, $\hat{\theta}$ and the predetermined confidence interval $I_n(\hat{\theta})$ are computed. Finally the test ϕ_* is said to reject the null hypothesis for that data set if

$$n^{1/2}T_n > c_s^*(\hat{\theta},\alpha,n) \equiv \max\{c_s(\theta,\alpha,n) : \theta \in I_n(\hat{\theta})\} .$$

Note that because c_s is a cubic spline, this maximization is quite trivial. With this modification, the computer evaluation of ϕ_* can be

done very quickly.

To achieve maximum correlation in the results, the same simulated data sets were used to assess ϕ_{COX}. The standard errors in the resulting probabilities of rejection are roughly about .002.

Because ϕ_{PAR} does not permit a similar modification, its evaluation is included only in Example 4.1. There, for each of 10^3 sets of pseudo-random samples, 201 bootstrap samples were simulated to obtain $c'(\hat{\theta},\alpha,n)$. The standard error of the results for ϕ_{PAR} is therefore at least .008.

All the computations were done on a VAX 11/750 computer. Pseudo-random numbers were generated via the International Mathematical and Statistical Library, and the FORTRAN program in Forsythe, Malcolm and Moler (1977, Chap. 4) used to fit cubic splines.

Example 4.1. Lognormal versus Exponential.

Let $(X_1,\ldots,X_n)$ be a random sample from a distribution with density $h(x)$ and consider testing

$$H_0 : h(x) = f(x,\mu,\sigma) \quad \text{vs.} \quad H_1 : h(x) = g(x,b)$$

where $x > 0$, and

$$f(x,\mu,\sigma) = \{x\sigma(2\pi)^{1/2}\}^{-1} \exp\{-(\log x - \mu)^2/(2\sigma^2)\}$$

$$g(x,b) = b^{-1} \exp(-x/b) \quad .$$

Jackson (1968), Atkinson (1970) and Epps, Singleton and Pulley (1982) have also considered this problem. We will use ϕ_{ESP} to denote the latter test in the sequel. From Cox (1962), the nominal level α rejection region for ϕ_{COX} is

$$\hat{\mu} - \log \bar{X} + \hat{\sigma}^2/2 > n^{-1/2} z_{1-\alpha}\{\exp(\hat{\sigma}^2) - 1 - \hat{\sigma}^2 - \hat{\sigma}^4/2\}^{1/2} \qquad (4.1)$$

where $\hat{\sigma}^2 = n^{-1} \Sigma(\log X_i - \hat{\mu})^2$ and $\hat{\mu} = n^{-1} \Sigma \log X_i$. Also,

$$T_n = \log \hat{\sigma} + \hat{\mu} - \log \bar{X} + (\log (2\pi) - 1)/2 \quad .$$

To understand the relative merits of ϕ_{COX} and ϕ_*, the following remarks may be helpful. First note that the problem is invariant under scalar multiplication and both ϕ_{COX} and T_n are scale invariant. Therefore the problem can be reduced by restricting to scale invariant tests. It can be verified that for each σ_0, the uniformly most powerful invariant test of the smaller null hypothesis $H_0' : h(x) = f(x,\mu,\sigma_0)$ vs. H_1 rejects H_0' if

$$S_n(\sigma_0) \equiv n^{-1}(n-1) \log \sigma_0 + \hat{\mu} + \hat{\sigma}^2/(2\sigma_0^2) - \log \bar{X}$$

is too large. Since the LHS of (4.1) is equal to $S_n(1)$, we see that ϕ_{COX} is an approximation to the uniformly most powerful invariant test for $\sigma_0 = 1$. The value $\sigma = 1$ is in some sense least favorable because under H_0,

$$ET_n \to \log \sigma - \sigma^2/2 + \text{constant}, \quad \text{as} \quad n \to \infty \quad ,$$

and the limit is maximized at $\sigma = 1$. On the other hand, $T_n - S_n(\hat{\sigma}) \to .5 \log (2\pi) - 1$ as $n \to \infty$. Therefore, since ϕ_* and ϕ_{PAR} reject for large values of T_n, they also approximate the uniformly most powerful invariant test, but by first estimating the unknown σ_0 with $\hat{\sigma}$. These observations suggest that ϕ_{COX}, ϕ_{PAR} and ϕ_* would all be reasonable for large n.

We investigated the performance of these tests by Monte Carlo simulation for $n = 20$. Table 4.1 presents the results for the probability of a type I error, α_I. The figures for ϕ_{ESP} are quoted from Epps, Singleton and Pulley (1982). Two spline-fitted curves c_s

are used, one to obtain the first two rows of Table 4.1, and another for the remaining rows. The first curve is based on a grid of twenty-five equally spaced values of $\log \sigma$ centered at $\log(.007)$. The second curve, shown in Figure 4.1, uses a grid of thirty-one equally spaced $\log \sigma$ values centered at 0. The spacing of the grid points for both curves is $(2n)^{-1/2} \log \log n$. In the computations for ϕ_*, we took the interval with endpoints $\log \hat{\sigma} \pm (2/n)^{1/2} \operatorname{loglog} n$ as the confidence interval $I_n(\hat{\theta})$ for $\log \sigma$. Table 4.2 shows the powers of the four tests. The data indicates that besides having very low power, ϕ_{COX} has size in excess of 0.4. ϕ_* appears to control the significance level quite well, with only slight loss of power compared to ϕ_{PAR}.

Table 4.1. Prob (Type I error) for H_0 : lognormal vs. H_1 : exponential, $\alpha = .05$, $n = 20$.

σ	ϕ_{COX}	ϕ_{ESP}†	ϕ_{PAR}	ϕ_*
0.005	.4056	?	0	0
0.01	.1229	?	0	0
0.5	.0172	.066	0	0
1.0	.002	.066	.049	.0206
1.414	.0001	.066	.061	.0497
2.0	.0001	.066	.024	.0127

†From Epps, Singleton and Pulley (1982).

Fig. 4.1. Critical values for lognormal vs. exponential

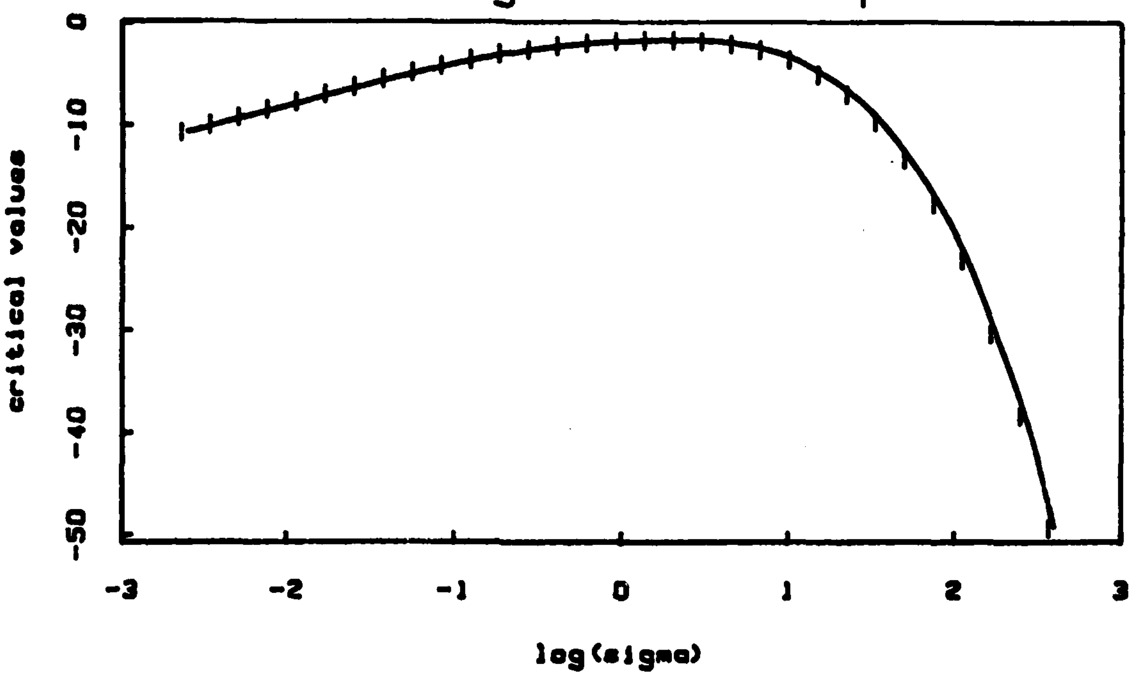


Table 4.2. Power of tests for H_0 : lognormal vs. H_1 : exponential, $\alpha = .05$, $n = 20$.

ϕ_{COX}	ϕ_{ESP} †	ϕ_{PAR}	ϕ_*
.0109	.35	.471	.4253

† From Epps, Singleton and Pulley (1982).

Tables 4.3 and 4.4 show the corresponding results with the roles of the hypotheses interchanged. Now ϕ_{PAR} and ϕ_* are the same tests because the distribution of T_n is invariant over H_0 : exponential. The powers of ϕ_{COX} and ϕ_* appear comparable.

Table 4.3. Prob (Type I error) for H_0 : exponential vs. H_1 : lognormal, $\alpha = .05$, $n = 20$.

ϕ_{COX}	ϕ_{ESP} †	ϕ_{PAR}, ϕ_*
.0587	.105	.0517

†From Epps, Singleton and Pulley (1982).

Table 4.4. Power of tests for H_0 : exponential vs. H_1 : lognormal, $\alpha = .05$, $n = 20$.

σ	ϕ_{COX}	ϕ_{ESP} †	ϕ_{PAR}, ϕ_*
0.5	.9999		.9999
1.0	.4014	.22	.3713
1.414	.5482	.60	.5297
2.0	.8951		.8885

†From Epps, Singleton and Pulley (1982).

Example 4.2. Poisson versus Geometric.

Let $(X_1,\ldots,X_n)$ be a random sample from a distribution with probability function $h(x)$. We wish to test $H_0 : h(x) = \lambda^x \exp(-\lambda)/x!$ vs. $H_1 : h(x) = \beta^x/(1+\beta)^{x+1}$, where $x = 0,1,2,\ldots$ in either case. The maximum likelihood estimators $\hat{\lambda}$ and $\hat{\beta}$ are both $\bar{X}$ and $T_n = n^{-1} \Sigma \log X_i! + \bar{X} - (1+\bar{X})\log(1+\bar{X})$. Cox (1962) showed that ϕ_{COX} has the rejection region

$$\Sigma \log X_i! - n\ell_f(\bar{X}) > z_{1-\alpha}\{nv_f(\bar{X})\}^{1/2}$$

where ℓ_f and v_f are functions defined therein. A short table of values of these as well as other needed functions are given in Cox (1962). We obtained other values by spline interpolation.

With $n = 20$, a grid of 20 $\lambda^{1/2}$-values was used to construct the spline-smoothed curve c_s. The confidence interval $I_n(\hat{\lambda}) = (\hat{\lambda}^{1/2} - n^{-1/2} \log\log n, \hat{\lambda}^{1/2} + n^{-1/2} \log\log n)$ $(0,\infty)$ for $\lambda^{1/2}$ was used in the computations for ϕ_*. Table 4.5 presents the results. Corresponding results with the hypotheses interchanged are shown in Table 4.6. Here the confidence interval for $\beta^{1/2}$ used was $I_n(\hat{\beta}) = (\hat{\beta}^{1/2} - n^{-1/2} \log\log n(1+\hat{\beta})^{-1/2}, \hat{\beta}^{1/2} + n^{-1/2} \log\log n\,(1+\hat{\beta})^{-1/2})$ $\cap\ (0,\infty)$. It is clear from the tables that, for the parameter values considered, ϕ_{COX} and ϕ_* are practically equal in performance. For values of λ and β closer to 0, however, the discrete nature of T_n will progressively cause ϕ_{COX} and ϕ_* to have arbitrarily low power, and some sort of randomization will be necessary.

Table 4.5. Prob (Type I error) and power for H_0 : Poisson vs. H_1 : Geometric, $\alpha = .05$, $n = 20$.

λ	Prob (Type I error) ϕ_{COX}	ϕ_*	β	Power ϕ_{COX}	ϕ_*
.30	.0527	.0309	.30	.199	.156
.45	.0488	.0452	.45	.272	.261
.60	.0478	.0475	.60	.362	.360
.75	.0431	.0427	.75	.453	.450
.90	.0427	.0398	.90	.543	.533

Table 4.6. Prob (Type I error) and power for H_0 : Geometric vs. H_1 : Poisson, $\alpha = .05$, $n = 20$.

β	Prob (Type I error) ϕ_{COX}	ϕ_*	λ	Power ϕ_{COX}	ϕ_*
.30	.0039	.0040	.30	.0185	.0190
.45	.0111	.0137	.45	.0800	.0895
.60	.0182	.0250	.60	.165	.201
.75	.0200	.0315	.75	.254	.321
.90	.0241	.0397	.90	.351	.435

Example 4.3. Quantal response.

Let $(X_1,\ldots,X_k)$ be independently binomially distributed with indices $n_1,\ldots,n_k$ and parameters $f_1(\gamma),\ldots,f_k(\gamma)$ under H_0 and $g_1(\beta),\ldots,g_k(\beta)$ under H_1, where $f_j(\gamma) = 1 - \exp(-\gamma x_j)$ and $g_j(\beta) = 1 - \exp(-\beta x_j) - \beta x_j \exp(-\beta x_j)$ for a set of "dose levels" $x_1,\ldots,x_k$. H_0 and H_1 have been called the "one-hit" and "two-hit" hypotheses

respectively. Following the example in Cox (1962), we chose 5 dose levels with $x_1 = .5$, $x_2 = 1$, $x_3 = 2$, $x_4 = 4$ and $x_5 = 8$. Also, we took $n_i = 30$, $i = 1,\ldots,5$.

The maximum likelihood estimators $\hat{\gamma}$ and $\hat{\beta}$ were computed iteratively using the algorithm in Thomas (1972). In the computation for ϕ_* we used as a confidence interval for γ the intersection with $(0,\infty)$ of the interval with endpoints $\hat{\gamma} \pm 2 \log\log n(ni(\hat{\gamma}))^{-1/2}$, where $i(\gamma) = \Sigma f_j'(\gamma)^2/\{f_j(\gamma)(1 - f_j(\gamma))\}$ is the information for γ. Table 4.7 presents the results of the simulation. Corresponding results with the hypotheses interchanged are shown in Table 4.8. As in the preceding example, ϕ_{COX} and ϕ_* appear quite comparable, with the latter keeping the significance level slightly better.

Table 4.7. Prob (Type I error) and power for H_0 : One-hit vs. H_1 : Two-hit, $\alpha = .05$, $n_i = 30$.

γ	Prob (Type I error)		β	Power	
	ϕ_{COX}	ϕ_*		ϕ_{COX}	ϕ_*
.05	.0406	.0474	.10	.252	.327
.10	.0465	.0444	.50	.897	.890
.50	.0436	.0414	1.0	.839	.830
1.0	.0440	.0477	1.5	.704	.696

Table 4.8. Prob (Type I error) and power for H_0 : Two-hit vs. H_1 : One-hit, α = .05, n_i = 30.

β	Prob (Type I error)		γ	Power	
	ϕ_{COX}	ϕ_*		ϕ_{COX}	ϕ_*
.10	.0651	.0478	.05	.724	.677
.50	.0535	.0418	.10	.856	.815
1.0	.0519	.0411	.50	.848	.827
1.5	.0461	.0394	1.0	.645	.634

5. Concluding remarks.

We have proposed here a test of separate families of hypotheses which requires very different assumptions from those for tests based on asymptotic normality of the test statistics, and showed in a series of examples that it is quite reasonable. The following points however should be mentioned. First, it is obvious that the results in section 2 remain true if any consistent estimator $\hat{\theta}$ is used instead of the maximum likelihood estimator. Further, although these results do not require conditions on the estimator $\hat{\omega}$ in (1.1), it is intuitively plausible that if high power is to be achieved for ϕ_*, $\hat{\omega}$ should at least be consistent. This condition is satisfied in all the examples.

It will be noticed that in the examples, we always have the critical value $c(\theta,\alpha,n)$ such that it is either independent of θ or a function of a one-dimensional component of θ. In situations where this is not so, the practical implementation of ϕ_* can be difficult, since we have to search for the maximum of a function in high dimensional space. In contrast, ϕ_{COX} does not have the same computational problem. However, as we saw in Example 4.1, the size of ϕ_{COX} may not be close to its nominal level, and this phenomenon may worsen in higher dimensions.

Acknowledgement

The author is grateful to Professor E. L. Lehmann for introducing him to the problem.

References

1. Atkinson, A. C. (1970). A method for discriminating between models (with discussion). J. R. Statist. Soc. B 32, 325-353.

2. Bahadur, R. R. (1971). Some limit theorems in statistics. Philadelphia: Society for Industrial and Applied Mathematics.

3. Brown, L. D. (1971). Non-local asymptotic optimality of appropriate likelihood ratio tests. Ann. Math. Statist. 42, 1206-1240.

4. Cox, D. R. (1961). Tests of separate families of hypotheses. Proc. 4th Berkeley Symp. 1, 105-123.

5. Cox, D. R. (1962). Further results on tests of separate families of hypotheses. J. R. Statist. Soc. B 24, 406-423.

6. Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. Philadelphia: Society for Industrial and Applied Mathematics.

7. Epps, T. W., Singleton, K. J. and Pulley, L. B. (1982). A test of separate families of distributions based on the empirical moment generating function. Biometrika 69, 391-399.

8. Forsythe, G. E., Malcolm, M. A. and Moler, C. B. (1977). Computer Methods for Mathematical Computations. Englewood Cliffs, NJ: Prentice-Hall.

9. Jackson, O. A. Y. (1968). Some results on tests of separate families of hypotheses. Biometrika 55, 355-363.

10. Pereira, B. de B. (1977). A note on the consistency and on the finite sample comparisons of some tests of separate families of hypotheses. Biometrika 64, 109-113.

11. Sawyer, K. R. (1983). Testing separate families of hypotheses: an information criterion. J. R. Statist. Soc. B 45, 89-99.

12. Shen, S. M. (1982). A method for discriminating between models describing compositional data. Biometrika 69, 587-595.

13. Thomas, D. G. (1972). Tests of fit for a one-hit vs. two-hit curve. Appl. Statist. 21, 103-112.

14. White, H. (1982). Regularity conditions for Cox's test of non-nested hypotheses. J. Econometrics 19, 301-318.

15. Wilde, D. J. (1964). Optimum seeking methods. Englewood Cliffs, NJ: Prentice-Hall.

16. Williams, D. A. (1970a). Discrimination between regression models to determine the pattern of enzyme synthesis in synchronous cell cultures. Biometrics 28, 23-32.

17. Williams, D. A. (1970b). Discussion of Atkinson (1970): A method for discriminating between models. J. R. Statist. Soc. B 32, 350.

WYL/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2574	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Parametric Bootstrap Procedure for Testing Separate Families of Hypotheses		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Wei-Yin Loh		8. CONTRACT OR GRANT NUMBER(s) MCS-7927062, Mod. 2 DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics & Probability
11. CONTROLLING OFFICE NAME AND ADDRESS See Item 18 below		12. REPORT DATE September 1983
		13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Parametric bootstrap; Separate families of hypothesis; Spline; Golden-section search.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is shown that the size of the parametric bootstrap procedure of Williams (1970) is biased upward. A bias-corrected version is shown to be better. The finite-sample performance of this procedure is examined and compared with that of Cox's (1961, 1962) test in a number of examples. In some of these, the proposed test reduces to the traditional optimal test.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

11-83

DTIC